DOCUMENT RESUME

ED 477 044 IR 021 772

AUTHOR	Kmiec, Angela; Pinchback, Melissa Piper; Rebelsky, Samuel A.
TITLE	Summarizing Links: Issues and Interfaces.
PUB DATE	2002-06-00
NOTE	7p.; In: ED-MEDIA 2002 World Conference on Educational Multimedia, Hypermedia & Telecommunications. Proceedings (14th, Denver, Colorado, June 24-29, 2002); see IR 021 687.
AVAILABLE FROM	Association for the Advancement of Computing in Education (AACE), P.O. Box 3728, Norfolk, VA 23514. Tel: 757-623-7588; e-mail: info@aace.org; Web site: http://www.aace.org/DL/.
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	Access to Information; *Computer Interfaces; Computer System Design; Hypermedia; Internet; Screen Design (Computers); User Needs (Information); World Wide Web
IDENTIFIERS	*Communication Links; *Links (Indexing)

ABSTRACT

At the center of hypertext is the link, an active connection from part of one document to another document (or part of that document). In the early days of the World Wide Web, it was easy for readers to identify the links on the page: the unvisited links appeared in blue and underlined, the visited links appeared in red or purple and underlined. With the advent of advanced formatting techniques, such as cascading style sheets, links appear in different ways on different pages. These changes may make pages appear more appealing, but they also make it harder for readers to engage in a key hypertext reading activity, identifying and selecting the most relevant links. At the same time, even easily identifiable links provide readers with very little accompanying information as to why the link is there and where it leads. This paper describes ongoing research in providing readers with additional access to links and information about those links. It revisits in more detail the need to provide links in a consistent and clear manner, considers ways to make links more useful to readers without affecting the design of a page, suggests techniques for adding information to links, and describes a tool the authors built that permits readers to obtain a summary of the links available on each page they visit. (Author)

Reproductions supplied by EDRS are the best that can be made
from the original document.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

G.H. Marks

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Summarizing Links: Issues and Interfaces

Angela Kmiec, Melissa Piper Pinchback, and Samuel A. Rebelsky
Grinnell Laboratory for Interactive Multimedia Experimentation and Research
Grinnell College, Grinnell, Iowa, USA
rebelsky@grinnell.edu

ED 477 044

Abstract: At the center of hypertext is the *link*, an active connection from part of one document to another document (or part of that document). In the early days of the World Wide Web, it was easy for readers to identify the links on the page: the unvisited links appeared in blue and underlined, the visited links appeared in red or purple and underlined. With the advent of advanced formatting techniques, such as cascading style sheets, links appear in different ways on different pages. These changes may make pages appear more appealing, but they also make it harder for readers to engage in a key hypertext reading activity, identifying and selecting the most relevant links. At the same time, even easily identifiable links provide readers with very little accompanying information as to why the link is there and where it leads.

In this paper, we describe ongoing research in providing readers with additional access to links and information about those links. We revisit in more detail the need to provide links in a consistent and clear manner, consider ways to make links more useful to readers without affecting the design of a page, suggest techniques for adding information to links, and describe a tool we have built that permits readers to obtain a summary of the links available on each page they visit.

1 Introduction

Where can I go from here? That question forms the center of hypertextual reading. Unlike in traditional linear writing, in which readers simply follow the narrative from beginning to end, in hypertext readers regularly look for links and consider whether or not to follow them. To permit readers to select the links that are most applicable, a hypertext system should provide four key attributes for links, summarized in Table 1.

Attribute	Summary
Identifiability	The links on a page must be easily identifiable.
Link Text	The source of the link must be obvious to readers.
Link Type	The purpose of the link should be clear.
Link Destination	The destination of the each link should be easy to determine.

Table 1: Key Attributes of Hypertext Links

1. ***Identifiability:*** **The links on a page must be easily identifiable.** That is, readers should be able to quickly tell what on a page is a link and what on a page is not a link. If readers cannot easily identify the links on the page, they will not follow the links. While some experimental hyperfictions may make the problem of finding the links part of the reading experience, most hypertexts should reveal, rather than obscure the links.
2. ***Link Text:*** **The source of the link should be obvious.** Most links do not go from page to page. Rather, they go from part of a page to another page (or part of another page). In a page with many links, readers should be able to clearly identify which part of the page the link is associated with. Typically, the source is a word, phrase, or image (the "anchor text" in HTML).
3. ***Link Type:*** **The purpose of each link should be clear.** A reader should know why the link is there. Does it provide a definition of the linked text? Some examples? A related page? A counter-argument to an argument on the page? The next page in a sequence? Without some additional information, readers are unlikely to know what the link is there for. It is this aspect of links that is perhaps least well supported on the Web.
4. ***Link Destination:*** **The destination of each link should be easy to determine.** Where does the link go? Is it elsewhere on the page, on the same site, or on another site? What kind of page does it

IR021772

BEST COPY AVAILABLE

link to? While many browsers will put a URL in the status bar, a URL only says a little about the destination.

These attributes are particularly desirable for educational hypertexts, since we put links in pages so that students can explore further as their desires or needs for further information guide them. By making links easily identifiable, we let students find information more quickly. By telling them why we put the link there, we help them decide whether it's a kind of link they need to follow. And, by telling them what kind of information they will receive, we help them decide whether or not that information is appropriate before they spend their time following and exploring a link. Note that the last two goals, while similar, are different. For example, we might provide a link for students who do not understand a short reading and need more information (that is, the purpose of the link). But a "need more information" link might lead to another reading, to a problem set, or even to a set of dictionary entries. By providing both kinds of information, we further guide the student.

Unfortunately, as we suggest in Section 2, Web pages are often incomplete in their support of these four key aspects. Designers are making links harder to identify. Page authors too often use "Click here" as their source text. Browsers provide little support for link types. And the primary information most readers get on the destination of a link is little more than the URL for the link. Even if the designer of an educational hypertext is careful to clarify links, choose good link text, type links, and provide information on destinations, students may not be able to access all this information and, in any case, lose these benefits as soon as they leave the site.

So, what can be done? In this paper, we propose a partial solution. *We have developed a prototype system that provides a link summary at the end of each page that readers view, no matter where on the Web those pages are located.* For each link on the page, the summary includes the source text, the full URL of the destination, the link type (if available), and additional information about the link or destination (if available). In the near future, we expect to add summary information about the destination page (its title, its size, etc.). Because the links appear at the end of the page, they are easy to identify. In pages that provide the appropriate accompanying information, the type of the link and basic information on the destination are available. Hence, the summary helps meet most of the criteria required for the links without affecting the design of the page.

In Section 2 of this paper, we revisit the current status of links on the World Wide Web and describe some deficiencies of current practices. In Section 3, we discuss reasons to provide link summaries on pages. In Section 4, we suggest key aspects that any link summary system should provide. In Section 5, we describe the architecture of our link summary system. In Section 6, we revisit the need to summarize and clarify links. Finally, in Section 7, we describe planned updates to our system.

2 Links and the World Wide Web

The World Wide Web (Berners-Lee et al. 1994) is the leading hypertext system. How good is its support for links? Unfortunately, the answer is "it depends", particularly on the page designer. In the early days of the Web, readers found it easy to identify links: Blue underlined text represented unvisited links, red or purple underlined text represented recently visited links. However, as the Web has evolved to provide authors with more control over the appearance of the page, each page designer has chosen new ways to show links. Readers can no longer easily identify the links on many pages by simply scanning for blue, red, and purple underlined texts.

Jakob Nielsen, one of the leaders in Web design, has emphasized the need for easily identifiable links in a number of his Alertbox columns. He first warned of the problem in 1996.

> Links to pages that have not been seen by the user are blue; links to previously seen pages are purple or red. Don't mess with these colors since the ability to understand what links have been followed is one of the few navigational aides that is standard in most web browsers. Consistency is key to teaching users what the link colors mean. (Nielsen 1996, Item 8)

Nielsen returned to this issue in a 1999 column. As new designs were becoming standard for Web sites, he repeated his warning. He also noted the importance of the underline as a way of identifying links.

> [Non-standard link colors continue] to be a problem since users rely on the link colors to understand what parts of the site they have visited. I often see users bounce repeatedly among a small set of pages, not knowing that they are going back to the same page again and again. (Also, because non-standard link colors are unpleasantly frequent, users are now getting confused by any underlining of text that is not a link.) (Nielsen 1999)

Designers can, of course, continue to use these standards, but fewer and fewer seem to. Readers must now spend more effort determining whether a word or phrase on a page serves as a link, often by considering the placement of a word or by waving the mouse over the word to see if anything happens.

So, the clarity and text of links were originally quite good on the Web, but have been weakened by the advent of careless design. What about the two other desired characteristics of links: their types and their destinations? Once again, the answer is "it depends". In part, it depends on the page author. In particular, the current HTML specification (Raggett et al., 1999) provides at least four attributes for every links: `title`, `rel` (relationship), `rev` (reverse relationship, from linked page to current page), and `class`. The title is a generic form of additional information that may be displayed by the browser. It can provide information about the destination. The relation and reverse relation tags allow authors to specify a link type. Unfortunately, the standard link types (Raggett et al. 1999, Section 6.12) are relatively basic, reflecting simple relationships like "table of contents" or "next element of a sequence" rather than more complex elements like those suggested in (Trigg 1983), such as "abstraction", "refutation", "continuation", "data". Particularly for educational texts, which will need types like "example", "exercises", "simplification", the basic set does not suffice. Finally, the `class` attribute exists primarily for formatting, but could also be used to indicate type.

Even more importantly, support for types and destinations depends not only on author, but also on browsers. For example, not all browsers support the important title attribute (e.g., Netscape 4.7 does not) and those that do often require readers to perform specific actions (e.g., pause the cursor for at least a few seconds) to obtain the title text. Few browsers seem to do anything useful with `rel` and `rev` attributes. And, while, many browsers support the `class` attribute, they do so for formatting rather than logical typing.

So, what should page authors do? They can eschew complicated site designs that obscure the links. They can choose careful texts for their links. They can use the `title`, `rel`, and `rev` attributes. But they still need to hope that browsers will support all that they do and that their readers will choose to use those browsers. Is that enough? We think not. By providing link summaries for all pages, we can support careful authors without affecting their careful designs.

3 Why Summarize Links?

In the introduction, we suggested that a link summary system provides one key characteristic that one hopes for in a hypertext system: it makes the links clear (they all appear the same way on every page and in the same place on every page). Since many pages obscure their links, we hope that provides reason enough for link summaries. However, the link summary system provides additional characteristics that readers benefit from: it automatically provides information about the destination of the link and, if the page author has included that information, can include the link type and other comments.

But there are still other reasons to summarize links at the end of a page. One of the benefits we first observed when using the system is that it aids readers in dealing with mistyped links in Web pages. Since the URL appears in the summary, it is possible to look at the URL and identify potentially incorrect parts of the URL. While this is a minor benefit, it is one that many users have found useful.

More importantly, a link summary system can permit readers to organize links. For example, on a page that presents an argument, one might want to follow all the links to counter-arguments. By arranging the link summary by link type, readers can quickly find all the counter-arguments. Similarly, a student reading a long reference piece might want to be able to quickly identify all the examples. Both cases require that authors classify their links, but we hope that as the types of links are made available through tools like ours, authors will be more inclined to classify their links. A reader might also arrange links by destination (e.g., to see whether there is a site or page that is linked to particularly often).

Should these summarized links replace the links on the page? Certainly not. Many links are best understood within a broader context. The link summary system simply provides a way for readers to get additional information on the links available in the page.

4 Design of a Link Summary System

Given the benefits derived from summarizing the links in a page at the end of a page, one would hope to see a number of systems designed to provide such summaries. In considering such systems, we should look for systems that meet a number of key goals that are summarized in Table 2.

Goal	Summary
Universality	Works with any page on the Web.
Browser Independence	Works with any reasonable browser.
Design Preservation	Maintains the original design of the page as much as possible.
Author Support	Takes advantage of information the author provides.
Customizability	Allows readers to set appropriate preferences.

Table 2: Primary Goals of a Link Summary System

1. Universality. The link summary system should allow readers to obtain summaries of any page on the World Wide Web without requiring modifications to that page. While we could hope that authors would provide their own summaries or use server-side software to provide summaries, not all authors will do so. A client-side system that works with any reasonable Web page provides a more universal system. Since a number of pages do not meet HTML standards, the system should do its best even with incorrect HTML.

2. Browser Independence. The link summary system should work with any browser, past, present, or future. Since different readers clearly prefer to use different browsers, the system should support as many browsers as possible. This requirement suggests that a browser plugin is an inappropriate implementation, since each browser will require a different kind of plugin.

3. Design Preservation. While many of the problems described earlier are caused by designers who ignore standards for link appearance, it is also true that many good designs may require links to appear differently. Hence, the link summary system should not modify the original page, except to add the links at the end (or in a separate window).

4. Author Support Given that the HTML standard provides a number of attributes for tags that can provide link type and destination information, the system should support those standards. Hence, it should take link types from the rel, rev, and class attributes and additional information from the title attribute.

5. Customizability. As the previous section suggests, different readers will wish to organize their links in different ways. The system should give readers appropriate control over the link summaries. Aspects of summaries that readers may wish to control include: the order of links within the summary (do they appear in the same order as on the page, or ordered by some other characteristics), the information associated with each link (just the source text and URL, or with additional information), and even how they appear (font, etc.).

5 Technical Details

We have developed a prototype system that supports the design goals suggested in the previous section. To provide a relatively universal system that is independent of browser, we rely on the Web Raveler architecture (Kensler and Rebelsky 2000). Web Raveler is a collection of systems that mediate the conversation between browsers and servers. The classic implementation of Web Raveler is as a proxy server. Proxy servers receive each request and response that is sent between client and server. Proxy servers are used for a variety of reasons, including caching of pages, limiting access to certain pages, and logging information about Web use.

Web Raveler provides an infrastructure in which authors can write small "plugins" (to Web Raveler) that receive each page before it goes to the browser and may modify the page as they choose. The Link Summary plugin in our prototype system quickly scans the page for links, extracts information, and adds a table of links at the end of the page.

Web Raveler also provides additional facilities useful for the link summary system. In particular, it provides account information which permits us to store and access user preferences on the appearance of the link summary. In addition, the Web Raveler team is currently developing a Web cache that will make it easy for us to quickly obtain the titles associated with many URLs. While it would possible to obtain the title of a Web page by sending an HTTP request, if there are many links on a page, it would be overly time consuming to send

an HTTP request for each of them. The Web Raveler cache would provide quicker access and user preferences could determine what the system does if the URL is not in the cache.

Our prototype system does a fairly simple search for links on a page. In particular, it looks for all the anchor (A) tags in the page. While this strategy works for most pages, it does not work for pages in which the links are generated on the client by Javascript or other scripting language. We are currently considering mechanisms for extracting such links and debating the tradeoffs (for example, interpreting a script will take additional time, making the system slower).

6 Conclusions

While links make hypertext hypertext, they are increasingly difficult to identify in many modern Web designs. In addition, the HTML standards that permit richer links are insufficiently supported by browsers. In this paper, we have suggested that a relatively simple mechanism, a link summary automatically generated for each page viewed, can increase the usability of modern Web pages and better support both readers and authors.

7 Future Work

As we suggest above, we hope to extend the system for provide additional information about the destination of each link. In its simplest form, this additional information can include the title of the page and the size of the destination page. However, it might also be useful to provide additional information about the destination, such as the author (if available) and even a summary.

Our prototype system was built as a basic proof of concept. In the near future we plan to begin some more careful user testing to see whether readers take advantage of link summaries and whether there are ways to make them more useful. We also hope to investigate what aspects of link summaries readers want to customize.

Because our link summary system appears to modify pages, we are also working to consider the intellectual property ramifications of the system. We are hoping to build on the work of the Web Raveler team in this direction.

References

Berners-Lee, T., Cailliau, R., Luotonen, A., Nielsen, H.F., and Secret, A. (1994). The World-Wide Web Communications of the ACM August 1994 Vol. 37, No. 8, pp. 76-82.

Fuller, G., Simonson, J., Tiwari, A., and Rebelsky, S. A. (2002). Clio's Assistants: A Tool Suite for Exploring Student Web Usage. Accepted to the EdMedia 2002 World Conference on Educational Multimedia, Hypermedia, and Telecommunications (June 24-29, Denver, Colorado).

Kensler, A. and Rebelsky, S. A. (2000). Web Raveler: An Infrastructure for Transforming Hypermedia. In J. Bourdeau and S. Heller (Eds.) *Proceedings of the EdMedia 2000 World Conference on Educational Multimedia, Hypermedia, and Telecommunications* (June 26-July 1, 2000, Montreal, Quebec, Canada), pp. 479-484. Charlottesville, VA: AACE, 2000.

Landow, George P. (1997). *Hypertext 2.0: The Convergence of Contemporary Critical Theory and Technology*. Baltimore, Maryland: The Johns Hopkins University Press.

Nielsen, J. (1996). Top Ten Mistakes of Web Design. Alertbox (monthly column), May 1996. Online document at <http://www.useit.com/alertbox/9605.html> (Visited 19 December 2001, Last modified 18 November 2001).

Nielsen, J. (1999). Using Link Titles to Help Users Predict Where They Are Going. Alertbox (monthly column), January 11, 1998. Online document at <http://www.useit.com/alertbox/980111.html> (Visited 19 December 2001, Last modified 25 August 2001).

Nielsen, J. (1999). "Top Ten Mistakes" Revisted Three Years Laters. Alertbox (monthly column), May 2, 1999. Online document at <http://www.useit.com/alertbox/990502.html> (Visited 19 December 2001, Last modified 25 August 2001).

Raggett, D., Le Hors, A., and Jacobs, I. (Eds.) (1999). HTML 4.01 Specification. Online document at <http://www.w3.org/TR/1999/REC-html401-19991224/> (Visited 19 December 2001, Last Modified 13 March 2000). World Wide Web Consortium (W3C) 24 December 1999.

Trigg, R. (1983). A Network-Based Approach to Text Handling for the Online Scientific Community. Ph.D. Thesis, University of Maryland. Chapter 4 available online at <http://www.workpractice.com/trigg/thesis-chap4.html> (Visited 19 December 2001, Last modified 13 July 2001).

Acknowledgements

We are particularly grateful to Grinnell College for the funding for the research described within this paper. The work benefited from the modern computing equipment in the department of Mathematics and Computer Science, equipment funded by Grinnell College and The Arthur Vining Davis foundation. Our work builds upon the prior work of others in the Grinnell Laboratory for Interactive Multimedia Experimentation and Research (Glimmer), particularly Andrew Kensler, who built the first version of Web Raveler, and the Summer 2000 Web Raveler team (Peter Broadwell, Elizabeth Hulse, Chris Kern, and Erin Nichols), who developed a more robust and modular implementation. We are also grateful to the other Summer 2001 research students who provided guidance in the Web Raveler system and who gave us suggestions for our own work: the Annotations team (Khong Lovan, Elias Vafiadis, and Joshua Vickery) and the Clio team (Greg Fuller, Joe Simonson, and Ananta Tiwari).

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

[X] This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

[ ] This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").